REPORT DOCUMENTATION PAGE

AFRL-SR-AR-TR-03-

0417

1. REPORT DATE *(DD-MM-YYYY)* 10-01-2003	**2. REPORT TYPE** Final Report	**3. DATES COVERED** *(From - To)* 06-01-1997 - 04-30-2003
4. TITLE AND SUBTITLE Understanding and Measuring Cognitive Workload: A Coordinated Multidisciplinary Approach		**5a. CONTRACT NUMBER**
		5b. GRANT NUMBER AFOSR #49620-97-1-0353
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) D. A. Boehm-Davis, W. D. Gray, L. Adelman, S. Marshall, R. Pozos		**5d. PROJECT NUMBER**
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) George Mason University, Psychology Department, MSN 3F5, Fairfax, VA 22030-4444		**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFOSR, 801 North Randolph Street, Room 732 Arlington, VA 22203-1977		**10. SPONSOR/MONITOR'S ACRONYM(S)** AFOSR
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approve for Public Release: Distribution Unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This research program was designed to develop predictive (based on cognitive modeling) and descriptive (based on physiological data) measures of cognitive workload that are highly correlated. Such measures must be theoretically grounded and empirically verified. Our main engineering goals in this project were to show (1) how the predictive measures (cognitive modeling) could be applied to guide the design of novel interfaces and communication protocols for decision making tasks, and (2) how the descriptive measures (physiological) could be used to measure workload during real-time task performance.

15. SUBJECT TERMS

cognitive modeling, ACT-R, pupil dilation, cognitive workload, EMG, control of attention, team performance

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** D. A. Boehm-Davis
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified		46	**19b. TELEPHONE NUMBER** *(include area code)* 703-993-8735

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

20031028 183

UNDERSTANDING AND MEASURING COGNITIVE WORKLOAD: A COORDINATED MULTIDISCIPLINARY APPROACH

Final Technical Report

Submitted by:

Deborah A. Boehm-Davis
Wayne D. Gray
Leonard Adelman

George Mason University

Sandra Marshall
Robert Pozos

San Diego State University

September 2003

UNDERSTANDING AND MEASURING COGNITIVE WORKLOAD: A COORDINATED MULTIDISCIPLINARY APPROACH

Final Technical Report
September 2003

Overview

This research program was designed to develop predictive (based upon cognitive modeling) and descriptive (based upon physiological data) measures of cognitive workload that are highly correlated. Such measures should be theoretically grounded and empirically verified. Our main engineering goals in this project were to show (1) how the predictive measures (cognitive modeling) can be applied to guide the design of novel interfaces and communication protocols for decision making tasks, and (2) how the descriptive measures (physiological) may be used to measure workload during real-time task performance.

Research Activities
GMU

The Argus Simulated Task Environment

The GMU side of the project focused its attentions on building a complex, simulated task environment, Argus (Schoelles & Gray, 2001). As discussed below, Argus had two major interfaces: Team Argus and Argus Prime. In both of its manifestations, Argus provided a task environment in which we could study a mix of cognitive, perceptual, and action operations that would be characteristic of the mix required by operators of systems such as AWACS, Patriot Air Defense, or other radar-monitoring tasks.

Argus was designed after an extensive investigation of similar simulated task environments. The systems investigated include Space Fortress (Donchin, 1995), Advanced Cockpit (Ballas, Heitmeyer, & Perez, 1992), the Team Interactive Decision Exercise for Teams Incorporating Distributed Expertise (TIDE2) (Hollenbeck et al., 1995; Hollenbeck et al., 1997), and Tandem (Dwyer, Hall, Volpe, & Cannon-Bowers, 1992). Like the Advanced Cockpit and Space Fortress, Argus places a premium on embodied cognition (Kieras & Meyer, 1997) and rapid shifts in serial attention (Altmann & Gray, 2002). Like TIDE2 and Tandem, Argus emphasizes judgment and decision making in a multiple-cue probability task (see also, Gilliland & Landis, 1992). Argus was designed to facilitate the investigation of a broad category of research questions centered on how interface design affects cognitive work load in both team and individual performance.

Beyond the simulation, Argus provides a suite of tools for creating task variations, manipulating experimental design, as well as data collection and analysis.

Detailed observations of human behavior in Argus was either the direct focus or the inspiration for the work performed at GMU. Many isssues of workload and interface were directly studied in the Argus Prime or Team Argus task environment. In other cases, issues arose in the study of Argus that could not be resolved in such a complex simulated task environment. These cases spun off a string of more controlled studies, the most notable and productive of which is the serial attention work.

Transfer of Technology

As of this writing Argus has successfully survived two technology transfers and may be poised for a third. Argus has gone from a tool being used solely at GMU to a second location; namely, Rensselaer Polytechnic Institute. The Rensselaer effort is notable in that Argus is being used there in a new line of research sponsored by an AFOSR grant. Argus was developed over a six-year period and the code badly needed updating. Some of the capabilities built into Argus had never been used, other of the capabilities had never been completely integrated. The move to Rensselaer Polytechnic Institute resulted in a complete overhaul of the Argus Prime software. The software now runs under the unix-based operating system, Mac OS X. Finally, discussions are proceeding for the technology transfer of Argus to other groups. In late Sept 2003, Rensselaer Polytechnic Institute was visited by Dr. M. Matessa from NASA-Ames. Dr. Matessa received hands-on training on all aspects of Argus Prime including a detailed review of the code. We expect that NASA-Ames will be using Argus Prime in future research projects.

Strategic Control of Attention

A notable feature of operator use of Argus Prime was that Argus required rapid shifts in attention – approximately 6-12 shifts per minute. As we were unclear as to how to model such shifts, we looked to the extant literature on task-switching (reviews of this literature are included in many of the Altmann and Gray publications listed in the appendix to this report.) At the time we began our review, the dominant accounts of task-switching postulated specialized cognitive components. These components were not specified at a mechanistic level and it was unclear to us whether the proponents of these approaches really believed that their components were structural as opposed to functional. Our theory holds that task-switching results from the dynamic organization of more basic cognitive components. This line of work has been carried from GMU to Michigan State University by Erik Altmann who was a post-doc at GMU when the project began. In addition to the many papers and presentations listed in the appendix, that work has resulted in many journal and conference publications since Erik left GMU for MSU (Altmann, 2002, 2003a, 2003b, in press-a, in press-b, in press-c; Altmann & Schunn, 2002; Altmann & Trafton, 2002; Trafton, Altmann, Brock, & Mintz, 2003). In addition, a manuscript with Gray is currently under review.

Period 1

The first period focused on the strategic control of visual attention. In the first study conducted, participants were interrupted during the execution of a task and instructed either to continue working on the same task or switch to a second task. Overall, the data showed that participants were slower to respond (by approximately 80 msec) on the first

trial after the interruption, even if they continued working on the same task. However, not all participants showed this deficit and several distinct patterns emerged, which tended to be stable within subject. We proposed that these differences in performance might reflect differences in "micro-strategies" as subjects try to find one strategy that will encompass both tasks. The research has led us to postulate dynamic micro-strategies that affect cognitive workload.

Period 2

In period two we began analyzing phenomena in the control of internal attention. Our technique involved collecting reaction time data that is accurate to the millisecond and attempting to construct theoretical models of our data by computational cognitive modeling. Three studies in this series were conducted. In a sample paradigm where the trial stimuli were single-digit numbers, a block of trials begins with an instruction to perform a simple task (e.g., classify the trial stimuli as odd/even). This instruction is followed by a run of trials. At some point, the run is interrupted by another instruction trial. This second instruction may tell Ss to continue the same simple task or switch to a new task (e.g., classify the trial stimuli as high/low). Trials are then continued until a total of 20 classification trials per block have been presented.

This simple paradigm revealed a wealth of phenomena that we believe contribute to an increasing cognitive workload in simple, but repetitive tasks. The effects noted by others are that the first classification trial after either initial instructions (T+1) or after the interrupting instructions (I+1) is reliably slower than the next trial (T+2 or I+2). This is called the <u>interrupt cost</u>. When the interrupting instruction switches the Ss to a new task, then I+1 is even slower (the <u>switch cost</u>) but the effects of switching task are gone by I+2. We noted an additional effect that we termed <u>within run slowing</u> (this effect has since been documented in, Altmann & Gray, 2002).

<u>Within run slowing</u> refers to the fact that Ss become gradually but reliably slower from T+2 to the next instruction or from I+2 (the second trial after the interrupting instruction) to the end of the block. This within run slowing increases at the rate of about 5 msec per trial with a cumulative impact of approximately 75 msec over six trials (T+2 to T+7 or I+2 to I+7). This previously undocumented effect is present when participants must remember what task to perform (odd/even judgements vs. high/low judgements, with digit stimuli in both cases), but absent when the task is implied by the stimulus (odd/even judgements on digits vs. consonant/vowel judgements on letters). Our initial explanation was that the slowing arises from interference among memory traces for different trials governed by the same instruction (Altmann & Gray, 1998). Closer examination of the data and attempts to model it caused us to modify this stance. We now believe that within run slowing is a byproduct of the same processes that produce both the interrupt cost and the switch cost. Working the details of the theory and obtaining supporting data is now documented by the many journal publications that came out of this work.

During period 2, three complete empirical studies were conducted and analyzed and a fourth, pilot study, was conducted. We believe that the phenomena underlying these simple tasks are pervasive in many tasks that confront operators of electronic equipment.

Indeed, we believe that many of the microstrategies developed by operators of much more complex equipment represent strategies to deal with the attentional and memory deficits illustrated by these simple paradigms.

Period 3

During period 3 we realized that our theory of task-switching applied beyond our paradigm to highly-dynamic task environments. The main result is that when mental attention must shift among items every 5 to 10 seconds, performance is constrained by the relatively slow rate of forgetting in human memory. Because old items decay gradually rather than instantaneously, and because memory is noisy, dynamic task environments involve a massive potential for interference from old items. To maintain accurate performance under such circumstances, cognition must deploy encoding and retrieval strategies that resist interference. We developed closed-form and computational models of these processes, applied them to data from a laboratory serial-attention task and to data from a problem-solving task involving memory for goals. The models established a lower bound on the time needed to commit an item to memory to be reliably retrieved for the next few seconds, and relate encoding time to performance accuracy. The predicted encoding-time constraint has implications for the role of appropriate external cues to offload memory, and for the task tempo at which operators can be expected to maintain specified levels of accuracy.

Period 4

The main development in the control of attention research concerned task tempo. Task tempo is, generally, the rate at which the task environment changes, and we used our models to ask how changes in the frequency of task switching should affect performance when the operator is responsible for tracking such changes.

Our algebraic and simulation models predicted that increasing the task tempo should decrease memory-related performance measures, because the time available for old items to decay should decrease and thus increase proactive interference. These predictions were tested using a task switching laboratory paradigm, and were supported both in terms of response time and error measures.

Our models characterize memory overload in terms of quantitative parameters such as memory-updates per unit time, and their promise appears to lie in the development of engineering models of tempo effects. Such models could be used to assess the memory requirements of dynamic task environments much as tools like GOMS are now used to assess perceptual, cognitive, and motor requirements. The work also contributes to cognitive theory (the ACT cognitive architecture in particular) in that it begins to examine how the memory system adapts to changes in the rate of change of the environment. The work is a step toward integrating two tracks of the Argus project, suggesting that memory overload may be a low-level architectural trigger for higher-level changes in decision-making strategy.

Other inner-loop work investigated ACT's attentional coding and associative learning mechanisms, both of which subserve performance in dynamic task environments like Argus. Finally, a successful simulation of incidental memory for order was developed.

Period 5

Over the first several years of the grant, we engaged in "inner loop" research focused on the nature of the strategic control of attention. Empirical work was combined with computational modeling work to understand the nature of task switching. The work led to the development of functional decay theory. This theory proposes that decay and interference, historically viewed as competing accounts of forgetting, are instead functionally related. Specifically, the theory posits (1) that when an attribute must be updated frequently in memory, its current value decays to prevent interference with later values, and (2) the decay rate adapts to the rate of memory updates. Behavioral predictions of the theory were tested in a task-switching paradigm in which memory for the current task had to be updated every few seconds, hundreds of times. No new empirical work was done at GMU during this last reporting period. However, both Altmann and Gray continue this work by developing on a major statement of the theory that received favorable reviews and is currently being revised for resubmission. Beyond this, Almann has continued the empirical work at MSU as documented by the many published and in-press articles that we cited at the beginning of this section.

Interface design and cognitive modeling

Period 1

Work within the Argus Prime simulated task environment suggested a line of research, which focused on cognitive modeling of interleaved tasks; that is, the execution of two or more individual tasks that can be performed in isolation or together. Although modeling individual tasks does not place new demands on our understanding of cognitive task analysis (CTA), modeling the concurrent execution of two tasks, or the rapid alternation of two simple tasks, does. First, the concurrent execution of tasks A and B may result in the creation of a new task, task AB. The CTA for task AB may be qualitatively different than what would be expected from a simple (or not so simple) interleaving of the elements of task A with those of task B. Second, the rapid alternation (e.g. tasks A B A B B A etc.) of two or more simple tasks may lead the user to perform each task differently than s/he would perform either task in isolation. As for the first case, the CTA of each task performed in isolation may be qualitatively different than that of the CTA of each task performed in alternation.

During the first period, in collaboration with SDSU, we began to examine the videotaped performance of expert users of the Aegis-based CIC task with the goal of modeling interactions with that system.

Period 2

Microstrategies develop in response to the fine-grained details of interface design. They are the users' way of optimizing interaction while minimizing the cost of that interaction. Microstrategies focus on what most designers would regard as the mundane aspects of

interface design – the ways in which subtle features of an interactive device interact with other aspects of an interface and task. However, although fine-grained, such details are important. This thread of Project Argus research is predicated on the assumption that milliseconds matter – 40 to 400 msec added to each routine of an interactive task may result in major workload problems in real-time, safety-critical systems or, less portentously, an interactive system whose interface simply feels soggy and awkward. (The work discussed here, eventually resulted in, Gray & Boehm-Davis, 2000)

In period 2, a major effort was the completion of the Argus simulated task environment – both Argus Prime and Team Argus. (These were discussed at the beginning of the GMU section of this report.)

During period 2, we developed the concept of microstrategies in the context of mouse clicks and mouse movements in a typical GUI interface. The development required some interesting theoretical (and practical) extensions to the CPM-GOMS cognitive task analysis technique. The technique was used to describe all available mouse move-click and click-move microstrategies. Gray and Boehm-Davis (2000) predicted that two different microstrategies would be used to click on buttons under two very slightly different context. The CPM-GOMS models of the microstrategies predicted a 150 msec difference in response times. The empirical data found a 136 msec difference. A follow on to the button study is ready to run in the fall. Our goals for this study were to determine how quickly and reliably microstrategies develop.

The Argus Prime part of the Argus Project entailed a search for cognitive components of workload. The quest was to identify low-level interface elements that can influence the performance of real-time, safety-critical tasks. The goal was to model the interaction of these components with human cognition during task performance using the computational cognitive modeling framework provided by ACT-R. Microstrategies is the intervening variable that we use to explain how low-level interface elements interact with a goal-driven cognitive architecture to produce differences in workload.

Period 3

In period 3, we continued to refine our concept of microstrategies and how they develop in response to the fine-grained details of interface design. Microstrategies are the users' way of optimizing interaction while minimizing the cost of that interaction. This thread of Project Argus research is predicated on the assumption that milliseconds matter – 40 to 400 msec added to each routine of an interactive task may result in major workload problems in real-time, safety-critical systems.

We conducted two experiments studying how microstrategies develop and contribute to workload using Argus Prime, a synthetic task that permits us to swap minor and/or major interface components while holding the task itself constant. Log files collected mouse clicks and point-of-gaze information to 17 msec accuracy.

The first experiment demonstrated that large differences in the interface (e.g., presenting information in a tabular versus a graphical format) influenced both overall performance

and the strategies used to accomplish the task. For example, the strategies used to select and acquire targets for classification were quite different for those participants using the tabular versus the radar display versions of the interface.

The second study examined the role of interface features in more depth using only the radar display interface. The data here indicated that strategies applied to target acquisition are sensitive to even small differences in interface design. Specifically, there was a reluctance to place information into working memory when external cueing was available as an alternative. This finding will be explored in more detail in the coming year. In these planned studies, design features of the interface will be manipulated to vary the amount of information that must be held in working memory. The impact on both performance and on cognitive workload will be assessed.

Another goal of this portion of the project was to model the interaction of interface components with human cognition during task performance using the computational cognitive modeling framework provided by ACT-R using microstrategies as an intervening variable. During the third period, preliminary ACT-R models were built using the perceptual-motor version of ACT-R (ACT-R/PM) to demonstrate that the models can interact directly with our software in a manner comparable to the ways in which our participants interacted with it.

During this period, we tried to connect our work to that being done at SDSU on physiological indicators of workload. Arguments have been made that eye blinks occur when cognitive processing of some stimulus is completed and that more complex processing should lead to a higher rate of blinks. We examined these hypotheses using data from the second Argus Prime study. We collected eye blinks from people performing this task to examine whether blinks occurred more frequently during periods of increased cognitive activities (during more complex scenarios) and as a cognitive punctuation mark (when a threat assessment is entered). The data supported the argument that blinks are associated with cognitive processing and that they may provide an initial indicator of cognitive workload.

Period 4

In the 4th period, we expanded the scope of our work in the area of interface design and cognitive modeling. First, we continued to collect empirical data on how microstrategies develop and contribute to workload using Argus Prime, a synthetic task that allowed us to swap minor and/or major interface components while holding the task itself constant. Log files collect mouse clicks and point-of-gaze information to 16.67 msec accuracy. In these studies, we continued to explore the impact of making subtle changes in the design of the interface on performance and on the strategies (and microstrategies) selected by participants. Specifically, in period 4 we manipulated the ease of retrieving history information from the display. Prior work on Argus had led us to postulate that the conditions in the task were such that subjects could have no memory for individual targets. On this assumption, an interface manipulation was made to create a condition where the participants should have performed extremely poorly. In fact they performed better than expected.

Second, we continued computational modeling of this task. Here again, our goal was to understand how subtle aspects of an interface might lead to large increases in cognitive workload. The modeling activity was based on the ACT-R/PM architecture, which combines ACT-R's theory of cognition with modal theories of attention and motor movement. This level of modeling allowed us to represent the microstrategies that we observed our participants using in the Argus Prime task into a computational cognitive model. The models demonstrated that interactive behavior in complex tasks is constrained not only by cognition but by perception and motor processes as well. Although these constraints exist at the millisecond level, the milliseconds added to a single interaction matter when the task requires thousands of interactions over an extended period of time. Further work on the modeling included expanding the modal models of visual attention and motor movement as well as working to incorporate a modal model of eye movements. These expansions are necessary to build models that respond adaptively to subtle differences in interface design. (Note that this work resulted in a doctoral thesis, Schoelles, 2002)

Third, we have developed an ACTion PROtocol analyzer (ACT-PRO). Discrete action protocols consist of time-stamped discrete user actions such as mouse clicks and key presses. Analysis of these action protocols often entails determining how well data match higher-level patterns (where those patterns are specified a priori by the researchers). Unfortunately, the process of sorting through thousands of actions to find matching patterns is very labor intensive. The action protocol analyzer that we have built provides two levels of pattern matching. Level one groups sequences of actions into sets of labeled strings. Level two matches these labeled strings to a hierarchical pattern. This allows us to use the tool to determine how well the data fit patterns specified by the experimenter. Complementarily, it can be used to focus the experimenter's attention on those data that do not fit the pre-specified patterns. (This work resulted in, Fu, 2001. This paper won the Castellan prize for best student paper at the annual meeting of the Society for Computers in Psychology.)

Period 5

In the last period of the project, we continued to collect empirical data on how microstrategies develop and contribute to workload using Argus Prime, a synthetic task that allows us to swap minor and/or major interface components while holding the task itself constant. Log files collect mouse clicks and point-of-gaze information to 16.67 msec accuracy. In these studies, we have continued to explore the impact of making subtle changes in the design of the interface on performance and on the strategies (and microstrategies) selected by participants. Specifically, this period we have manipulated the ease of retrieving history information from the display. Prior work on Argus had led us to postulate that the conditions in the task were such that subjects could have no memory for individual targets. On this assumption, an interface manipulation was made to create a condition where the participants should have performed extremely poorly. In fact they performed better than expected. We are now focusing on what strategies they used and how these strategies influenced workload.

Second, we have greatly expanded our work in computational modeling of this task. Here again, our goal is to understand how subtle aspects of an interface may lead to large increases in cognitive workload. The modeling activity is based on the ACT-R/PM architecture, which combines ACT-R's theory of cognition with modal theories of attention and motor movement. This level of modeling has allowed us to represent the microstrategies that we observed our participants using in the Argus Prime task into a computational cognitive model.

In this past period, the modeling work has specifically focused on making the model "embodied"; that is, the model now includes modal models of visual attention, motor movement, and eye movements.

Period 6

In the last period, we have continued to collect empirical data on how microstrategies develop and contribute to workload using Argus Prime, a synthetic task that allows us to swap minor and/or major interface components while holding the task itself constant. Log files collect mouse clicks and point-of-gaze information to 16.67 msec accuracy. In these studies, we have continued to explore the impact of making subtle changes in the design of the interface on performance and on the strategies (and microstrategies) selected by participants.

Second, we have continued to expand our work in computational modeling of this task. The modeling activity is based on the ACT-R/PM architecture, which combines ACT-R's theory of cognition with modal theories of attention and motor movement. In this past period, the modeling work has specifically focused on developing the "embodied" aspects of the model (i.e., modal models of visual attention, motor movement, and eye movements); we also exercised the model by running a number of model experiments, focusing first on how well the model could replicate individual subject data and then on models using different methods of target acquisition.

Dual Task Performance

Period 5

In the 5th period, we began to exercise the dual task aspect of the Argus Prime environment. The focus of much of our previous work on this grant concerned understanding the strategic control of attention and the impact of interface design decisions on the target classification task. The Argus Prime environment also allows for a dual task component, where the second task involves tracking a moving plane on one side of the screen. We have begun experiments where participants perform the tracking task while simultaneously performing the target classification task. The data will provide information on task switching at a higher level than that examined in our past work on strategic control of attention.

Period 6

In the past period, we have continued the work we began in the last reporting period to exercise the dual task aspect of the Argus Prime environment. The focus of much of our previous work on this grant concerned understanding the strategic control of attention

and the impact of interface design decisions on the target classification task. The Argus Prime environment also allows for a dual task component. We have run two series of experiments in which a second task was performed in addition to the classification task. In the first experiment, the second task involved tracking a moving plane on the right side of the screen. This task requires a high degree of visual and motor activity. In the second experiment, the second task is forced choice task in which a letter is spoken by the computer every four seconds and the participant is to respond via a key press whether the current letter is above or below the previous letter in the alphabet. The data will provide information on task switching at a higher level than that examined in our past work on strategic control of attention. In addition, the computational cognitive model has been extended to perform the tracking task in the dual task environment. (This work was recently reported in, Gray & Schoelles, 2003.)

Team decision making

Period 1

Our third area of focus was team decision making. Our initial effort in this area was directed toward (a) reviewing recently published literature, (b) designing an initial experiment that would examine the effects of time pressure on cognitive workload and team communication processes and performance, and (c) obtaining a laboratory task for that experiment. Thanks to Dr. Linda Elliott at the AESOP facility at Brooks AFB and personnel at Michigan State University (MSU), we obtained a current version of the $TIDE^2$ software used by Hollenbeck and Ilgen in their research on a multilevel theory of team decision making. However, our own experiences and those of other researchers not at Michigan State that we talked with suggested that this software was difficult to use. Therefore, we spoke with Dr. Stan Gully, a recent MSU graduate and assistant professor at GMU, about using the TANDEM software. However, in the final analysis, the entire research team decided to expedite the development of the Team module for the Argus system because it would most effectively permit testing of our hypotheses and integration among the various research thrusts of our MURI effort.

During the first project period, we prepared for and conducted an experiment examining how teams adapted to increasing levels of time pressure. Conceptually, the research was guided by Brunswikian theory, which focuses on trying to understand how individuals and teams adapt to different conditions in their environment. We used the multi-level, lens model that Brehmer and Hagafors developed in 1986 to extend Brunswik's lens model to the study of staff decision making, and that Hollenbeck and others have more recently used in developing their Multilevel Theory of Team Decision Making.

Operationally, we had 7 three-person teams participate in our study. Each team was composed of ROTC cadets, who participated for two hours per week for seven weeks. Our task was a dynamic, aircraft identification task using the Team Argus synthetic task developed during Period 1 at GMU. Two staff members (and a leader) had to track aircraft on their screens, pass information about the aircraft to each other, and make recommendations about the aircraft's level of hostility, which the leader could then use to make judgments while the aircraft were on the screen.

We made two principal hypotheses. First, we hypothesized that increased time pressure (i.e., less time to make a judgment about each aircraft), would lead to a reduction in the quality of the teams' decision making. Second, we hypothesized that teams would adapt (perhaps in different ways) in an effort to maintain decision quality. That is exactly what we found. Decision making quality decreased, although not as quickly or precipitously as predicted. In addition, there were few significant differences in the teams' overall performance scores. Teams did, however, adapt (or not) in different ways to increased time pressure. Specifically, three of the seven teams tried to continue performing the task as trained regardless of the time pressure; that is, the subordinates kept sending identification recommendations to the leader for all aircraft. In contrast, two teams simplified the task by having each subordinates make recommendations for only half the aircraft. And in two teams, the leader took over the entire decision making task by having subordinates only send information about the aircraft, not recommendations..

In addition, the leaders made a clear speed-accuracy trade-off in an effort to maintain performance. For example, in the condition with the greatest time pressure, the leader of one of the two leader-controlled teams made judgments for more aircraft than any other team, but had the lowest accuracy, which was defined as the correlation between the leader's decisions and the correct answer. In contrast, the leader for the other leader-controlled team had the highest accuracy score, but made the fewest number of judgments. Utilization of these (and other) adaptation strategies resulted in essentially equivalent levels of performance overall because none of the teams were able to maintain both speed and accuracy under high time pressure.

ACT-R model-building exercises were conducted to describe the differences in the adaptation strategies observed for teams in the experiment. These models were presented at the First-Year Annual Review meeting in May 1998, and at the Fifth Annual ACT-R Summer School at Carnegie Mellon University (Miller, 1998). In addition, the "team decision making" group developed a short description of their research as part of the project's larger submission to the Human Factors Conference, and presented their initial research findings at the Fourteenth Annual Meeting of the International Brunswik Society (Adelman, Henderson, & Miller, 1998).

Period 2

There were three major activities during Period 2. The first activity was completion of all data analysis for the first experiment. The analysis focused on trying to more fully understand how time pressure affected the participants' cognitive and communication processes and their adaptation strategies at both a micro-level (e.g., process acceleration) and a macro-level (e.g., different team processes), and decision making quality. The research results of the first experiment have been described in a brief book chapter (Adelman, Henderson, & Miller, 2001) and more fully, in a journal paper (Adelman, Miller, Henderson, & Schoelles, 2003).

The second activity was the development of a Hierarchical Decision Making (HDM) model for relating time pressure effects for dependent variables at different levels of granularity. The hierarchy had three principal levels of granularity. The top level

contained team-oriented dependent variables, such as team performance, the percentage of decisions made by the leader, and the leader's judgmental accuracy. The middle level contained subordinate-oriented dependent variables, such as staff validity and the percentage of recommendations made by each subordinate. And the bottom level contained interface-oriented variables, such as the number of times a target was examined. We used simultaneous multiple regression analysis to identify what lower-level variables affected the variables at the next level up the hierarchy over levels of the time pressure manipulation. The HDM model development and results were described in Henderson's (1999) dissertation.

The third activity was directed toward cognitively re-engineering the Team Argus interface to test implications of the statistical findings of the HDM Model, and observations made during the first experiment. Specifically, during Period 2, we prepared for and conducted an experiment testing the effectiveness of three different interfaces on team performance under increasing levels of time pressure. The first interface provided perceptual support by using colors to inform operators of the examination status of aircraft tracks, and symbols to inform them of when tracks where nearing decision points. It was hypothesized to maintain high levels of "percentage of decisions made" by addressing interface problems identified by the HDM Model. The second interface was the interface used in the first experiment. It provided a baseline against which to compare performance. And the third interface was the old system plus cognitive feedback. Cognitive feedback informed operators of their performance scores, their "percentage of decisions (and recommendations) made," and their judgment accuracy. In addition, it used multiple regression analysis to tell operators how they were weighting the cues and the extent to which the leader agreed with the recommendations of her/his subordinates. The value of cognitive feedback had been demonstrated with static tasks where, for example, there is only one aircraft on the radar display at a time. However, its value had not been tested in dynamic tasks like ours. Consequently, there was no empirical data assessing the relative importance of perceptual support versus cognitive feedback for team performance under high time pressure.

The second experiment showed, as hypothesized, that a perceptually-oriented interface could maintain an extremely high number of judgments as time pressure increased four-fold during the study. As a result, teams with perceptual support were able to maintain higher overall performance levels than those in the other two interface conditions. In fact, performance remained close to the training criterion even under time pressure levels that were four times greater than those used during training. Counter to our predictions, however, a cognitively-oriented interface providing feedback about how team members made their judgments did not maintain judgment accuracy as time pressure increased. Trying to understand why the cognitive feedback condition was not effective was a major activity of Period 3. The results of the second experiment were presented at the 44th Annual Meeting of the Human Factors and Ergonomics Society (Miller, Adelman, Henderson, Schoelles, & Yeo, 2000).

Period 3

There were three major activities in the team decision making area during Period 3. The first activity was additional analysis of the data from the second experiment. Specifically, during Period 3, a path model using lens model equation parameters and Multi-Level Theory constructs was developed to better understand the effect of time pressure on teams' judgmental accuracy. This analysis showed that the time pressure effect was fully mediated by decreasing team informity (amount of information held jointly across members). Leaders and their staffs stopped sending information as regularly; consequently, their decision making suffered. They were not able to use the judgment model they were trained to use, and independent of that, their judgments became less consistent. The cognitive feedback interface, which was developed to provide decision-making support, was unable to overcome this information breakdown. The problem was in keeping information flowing. The results of the path modeling effort were presented at the 16th Annual Meeting of the International Brunswik Society (Adelman, Yeo, and Miller, 2000). In addition, an invited book chapter discussing the details and importance of the modeling effort is currently in preparation (Adelman, Yeo, & Miller, in preparation).

As a result of the path model, the second principal activity focused on designing an experiment to assess whether adding simple enhancements to the perceptually-oriented interface could maintain information flow and judgment accuracy, in addition to judgment quantity, under even higher levels of time pressure than those used in the second experiment. The new experiment was with individual participants in a simulated team setting, an important advance made possible by the Argus system developed on the contract. This advance permitted us to disentangle the amount of information sent to each team member from the time pressure manipulation. This disentangling was important because the path model suggested that time pressure's effect on individual and team decision making was fully mediated by informity; that is, that time pressure had at best a minimal impact on individuals' decision making ability if they had the necessary information. Substantiation or rejection of this finding has important basic and applied research implications, particularly if we find, as predicted, that the interface is the overriding mediator of time pressure's effect on decision making.

The third major activity during Period 3 was beginning to perform a literature review investigating the effect of teammate interruptions on decision making performance. Past research has shed little light on the cognitive demands imposed by the different characteristics of interruptions in complex tasks, particularly the timing and relative importance of interruptions. What literature does exist focused on the costs or disruption associated with attending to an interruption. However, because interruption is such a frequent and non-trivial element of team communication, our research perspective focused on how individuals perceive and use interruptions to benefit their decision-making performance.

Period 4

In Period 4, we performed data collection and analysis for the third experiment outlined above. That experiment tested the effectiveness of a "Send" icon to support information flow and a "Receive" icon to support decision accuracy in a simulated distributed team

decision-making task varying time pressure, amount of information, and other task variables. As predicted, the "Send" icon was effective in maintaining information flow, particularly under high time pressure and when teammates tended to send less information, which is critical to maintaining the overall effectiveness of distributed teams. In contrast, the "Receive" icon was not effective, resulting in lower decision accuracy under the highest time pressure level. The decrement occurred because participants' using the "Receive" icon made a greater proportion of decisions with less information as time pressure increased, and with less cognitive control. This occurred because with increasing time pressure, participants adopted a strategy of making decisions before, not after, receiving information. Although unanticipated, the results illustrated the close and sometimes subtle relationship between the task, display, strategy, and performance.

The third experiment was important for three other reasons. First, in order to implement the experiment, project computer scientists modified Team Argus so it could simulate an actual team, an important advance over the earlier system. Second, the experiment controlled the information presented to participants. Consequently, we know that the decrease in decision accuracy caused by increasing time pressure was caused by a decrease in participants' cognitive control of the procedure that they were trained to use when making decisions, and not their knowledge of the procedure or adoption of a new procedure. Third, the experiment showed that participants with higher working memory capacity integrated more information, and that task variables (time pressure, amount of information, run number, scenario order, and type of information) had strong effects on behavior. The results of this experiment were presented at the 17th Annual Meeting of the International Brunswik Society (Adelman, Miller, & Yeo, 2001). A journal manuscript describing the third experiment has just been accepted for publication (Adelman, Miller, & Yeo, accepted).

In addition, Sheryl Miller completed her dissertation proposal on the effect of interruptions on team decision making during Period 4.

Period 5

In the third experiment, we found that an icon telling operators when they had received information about an aircraft did not improve their decision accuracy, and was particularly ineffective under the highest time pressure level. This result was surprising because we had predicted that decision accuracy would increase with the "Receive" icon because operators would adopt a strategy of waiting longer to gather more information before making a decision. However, operators adopted the alternative strategy of making decisions before, not after, receiving information, presumably in an effort to maintain the number of decisions they made.

During Period 5 we tested two different approaches for counteracting that strategy and, thereby, maintaining high levels of decision accuracy under the highest time pressure level. One approach was display-oriented; it involved placing a large white square in an appropriate location on the aircraft symbol whenever information arrived after operators had made a decision. This permitted operators to know when information arrived before

and after making a decision. The second approach was organizationally-oriented; it involved increasing the importance of making accurate decisions from 0.50 to 0.90, relative to making many decisions or sending a lot of information. This approach manipulated operators' reward structure by affecting the overall feedback score received at the end of each experimental session.

A factorial experiment was conducted with 2 levels of interface (old and new), 2 levels of reward structure (old and new), and 3 levels of time pressure (1.2 new aircraft on screen/sec, 2.4, and 3.6). Interface and reward structure were between-subject factors and time pressure was within subject. The experiment was conducted using the Team Argus system in the simulated team condition so that we could control the amount and timing of the information sent to each operator. We also gave participants the N-back working memory test because our third experiment had found working memory to correlate with decision accuracy scores on the Team Argus system. We used Analysis of Covariance to analyze the effect of interface, reward structure, and time pressure on decision accuracy.

We found that the "new" reward structure was extremely effective in maintaining high levels of decision accuracy, regardless of the level of time pressure. Consistent with our hypothesis, operators changed their strategy and waited longer and had more information when making decisions with the new reward structure. The cost, as predicted, was that they made fewer decisions and sent less information to their simulated teammates. In contrast, the interface had no affect on decision accuracy. Although it did foster re-decision making at lower time pressure levels, it failed to do so at the highest level. And, surprisingly, it had no affect when combined with the new reward structure. These results suggest that, depending on the team decision making task and support environment, (1) there is a time pressure level beyond which operators can not maintain both decision quantity and quality, and (2) if one wants them to maintain quality, the reward structure and not the interface, may be the more effective mechanism for making them do so. This research will be presented at the 2003 IEEE Systems, Man, and Cybernetics Society Conference, and published in its proceedings (Adelman & Gambill, 2003).

Interrupted Decision-Making

Periods 5 and 6

The Team Argus environment was modified during Period 5 to investigate the conditions under which messages were most effectively integrated with on- going decision-making tasks. This investigation focused on the effects of interrupting decision makers. Team Argus offered an interesting context in which to study interrupted decision-making, because it reflects many characteristics of real world interruptions. Interruptions are a frequent and expected part of decision-making, and they must be integrated into on-going taskwork. Four experiments were conducted using Team Argus.

The first experiment was designed to explore disruption as a consequence of the timing and relevance of interrupting messages. The Team Argus interface was modified so that incoming communications were composed of a) an alert and b) message data. Interruptions were unavoidable in that participants were unable to see the current

decision task once the alert appeared on the screen. Half of the participants were instructed to implement a memory strategy such that they actively tried to remember the task resumption point at the point of interruption. The other half of the participants received no such instruction. Analyses indicated that this strategy actually resulted in overall poorer time-on task, because participants using the memory strategy had difficulty balancing the needs to remember the task resumption point and to remember the content of the interrupting message.

The second and third experiments (differing only in terms of decision complexity) were used to investigate this balance. Two performance cost variables were manipulated, the cost of forgetting the task resumption point and the cost of forgetting the content of the interrupting message. Analyses indicated that these variables affect the decision processes in terms of the time spent switching attention from the primary decision to the interrupting message and the time spent actually attending to the message.

The fourth experiment used a further modified version of Team Argus, such that the interruption (alert and message data) did not prevent the participant from viewing the interrupted decision. The interruption was available for processing for 5 seconds. Thus, the participant could choose to read the message, delay reading the message, or entirely ignore the message. Analyses investigated the strategies that participants develop to process interruptions given the varying performance costs associated with different messages.

SDSU
Pupil Dilation and the Index of Cognitive Activity

Three primary goals of this project were:

- to create new psycho-physiological measures of cognitive workload,
- to demonstrate the reliability of these measures, and
- to determine whether they were suitable for measuring workload during real-time task performance.

These goals have been achieved with the Index of Cognitive Activity, a patented metric based on changes in pupil dilation.

Development of the Index of Cognitive Activity

Rationale

The predominant measure of changes in pupil dilation is Jackson Beatty's evoked pupil response created more than 20 years ago (Beatty, 1982). This technique is based on evoked response potentials used to measure event-related brain potentials in EEG. To apply the technique, researchers typically present a stimulus repeatedly at a fixed interval of time. A baseline recording is made prior to each stimulus presentation and the absolute change in pupil size is recorded several hundred milliseconds after presentation.

These recordings are then averaged across stimuli and across individuals to reach an estimate of how much the pupil responds to the particular task.

Beatty's method has proved very valuable in clinical applications, but it has severe limitations for practical applications. First, it requires a simple stimulus that can be presented repeatedly. Second, it measures average absolute changes in pupil size. And, third, it depends upon averaging across stimuli and individuals.

Practical assessment of cognitive workload requires the ability to measure sudden changes in pupil size as individuals engage in complex cognitive activities. For example, one might wish to assess the cognitive effort of a pilot landing a plane or a TAO directing the Combat Information Center of a ship. In such situations, the tasks are unique and non-repetitive. One or more critical events may occur at various points in time, and it is the response to these events that is of interest. The events do not occur at fixed intervals nor are they necessarily repetitions of the same crisis. They are unique events that emerge across a long time interval that must be measured continuously if pupil changes are to be accurately detected.

The pupil responds dramatically to changes in lighting, with the typical size for an individual ranging from about 2 mm to 8 mm when moving from bright light to darkness. Moreover, the eyes of individuals vary in size, with some people having larger pupils than others. Most studies of pupil changes using Beatty's technique have looked for average increases or differences of 0.1 mm, and changes of this size are usually statistically significant. However, such measurements require that individuals be screened to insure that they have similar-sized pupils initiall,y and they must be in well-controlled lighting conditions as well. Otherwise, the absolute size of the change is not meaningful, because a 0.1 mm change for a pupil that is 3 mm in diameter is quite different from a similar change in the same pupil that is 8 mm in diameter. Of higher utility is a metric that assesses relative increases in pupil size.

Finally, measures of cognitive workload must be valid for an individual. In applied settings, it is the single operator who will be assessed and whose workload must be measured in real time while he or she is performing the job. Metrics based on averaging—either across tasks or across individuals—are not sufficient.

In summary, a useful metric must be able to measure events across time, it should be a relative instrument that can be used in variable lighting, and it must measure a single user reliably. Each of these requirements is quite difficult to achieve.

<u>Technical Details</u>

The challenge in measuring pupil dilation is to separate two reflex responses that often occur simultaneously, the light reflex and the dilation reflex (Loewenfeld, 1993). The light reflex is the pupil's response to any light source, and it results in an irregular oscillation of the pupil through the process of reciprocal innervation. Two sets of muscles govern pupil dilation, the circular muscles surrounding the pupil and the radial muscles extending outward from the pupil. In the presence of light, the circular muscles

typically are activated while the radial muscles are inhibited, causing contraction of the pupil. In the presence of a cognitive stimulus, the radial muscles are activated and the circular muscles are inhibited, resulting in a burst of dilation larger than either muscle group could produce alone. The Index of Cognitive Activity (ICA) was developed to measure this dilation reflex.

The Index of Cognitive Activity is derived from wavelet analysis, using relatively recent developments in applied mathematics. Wavelet analysis consists of repeated orthogonal transformations of a signal. The goal is to decompose the original signal into several independent components, each of which can be analyzed and interpreted. At the heart of wavelet analysis is a 'mother wavelet,' a small oscillatory function that decays rapidly to zero in both positive and negative direction, i.e., a little wave. For a signal x and a mother wavelet ψ, the process of wavelet analysis is expressed by

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$$

where j is an index of dilation and k is an index of translation. Systematic variation of j and k will create a family of wavelets able to reproduce fully the original signal.

Wavelet analysis proceeds iteratively. Using the mother wavelet function, the dilation transformation first extracts the high frequency details from the signal by setting j=1 and evaluating all possible k. Next, using a scaling function that is orthogonal to the wavelet function, a second transformation extracts from the signal all information not captured by the wavelet transform. The initial wavelet transformation captures the largest abrupt changes or discontinuities in the signal. The scaling transformation results in a smoothing of the signal because these discontinuities have been removed from it.

The signal can be decomposed further if required by repeatedly applying the wavelet transformation (i.e., j=2, 3, …, for all k) and the associated scaling function to the result of the most recent scaling transformation. Thus, additional details of the signal are extracted with subsequent wavelet transforms, and the signal becomes smoother with each ensuing application of the scaling transform. The result of the full analysis is a smoothed approximation of the signal (obtained from the final scaling transformation) together with multiple sets of detail coefficients. All parts of this decomposition are orthogonal, and the original signal will be obtained if the last approximation and all sets of details are summed.

A key statistical question that arises in the analysis of signals such as the pupil dilation signal is whether significant change points can be identified. Mathematics researchers have shown that wavelets are well suited to solving statistical change-point problems when the objective is to determine whether the jumps observed in a signal are statistically significant. One first establishes a threshold and then sets all wavelet coefficients falling below the threshold to zero.

In the course of this project, a number of different wavelets and thresholds have been evaluated. Currently, the Daubechies wavelet of size 8 and a threshold of size 4 appear

to be the most satisfactory. Further research will be required to determine optimal values.

The Index of Cognitive Activity is computed from the results of the wavelet analysis after the threshold has been applied. The number of non-zero coefficients is tallied for each second of observation for the entire time. In many instances, it is useful to look at the average ICA across the entire time period, and this is achieved by tallying the total number of non-zero coefficients divided by the total number of seconds. If there are critical events to be measured, the location and duration of those events during the time period are identified, and the average ICA per second for those events may be computed as well. Thus, it is possible to compute the Index over the full task or to decompose the task into sub-tasks of any length and to examine them separately. Because the Index always reflects the same ratio—the frequency of occurrence per second—it provides a common basis for comparing individuals, groups of individuals, single events, and multiple events.

<u>Patent Information</u>
The process described above has been patented by the U.S. Office of Patents and Trademarks: *Method and Apparatus for Eye Trackinig and Monitoring Pupil Dilation to Evaluate Cognitive Activity.* Patent application approved February 2000, U. S. Patent No. 6,090,051.

Evaluation of the Index of Cognitive Activity

<u>Overview</u>
The Index of Cognitive Activity has been tested with a number of well-known psychological tasks as well as in several applied settings. Additionally, the procedure has been applied to data from different eye-tracking systems or pupillometers to ascertain that it is not system-dependent.

<u>Baseline Study</u>
The simplest validation of the Index comes from applying the procedure described above to data collected from a single individual under four conditions: (1) no pupil reflex, (2) light reflex only, (3) dilation reflex only, and (4) both reflexes simultaneously.

Four test conditions were designed. In each one, the individual looks for approximately 2 minutes directly at a computer screen placed about 18 inches away. The conditions are: do nothing while looking at a dark screen in a dark room, respond to verbal arithmetic problems while looking at a dark screen in a dark room, do nothing while looking at a lighted screen in a lighted room, and respond to verbal arithmetic problems while looking at a lighted screen in a lighted room.

The raw pupil data for one individual are presented in the four graphs of Figure 1, and the results of the analyses of these data are given in Figure 2. (Data in both figures have been normalized for comparative purposes.) As expected, the pupil signal is relatively calm in the 'dark plus no cognitive activity' in the upper left quadrant of Figure 1. The

presence of light results in an agitated signal (upper right and lower right) as does the presence of cognitive activity (lower left and lower right).

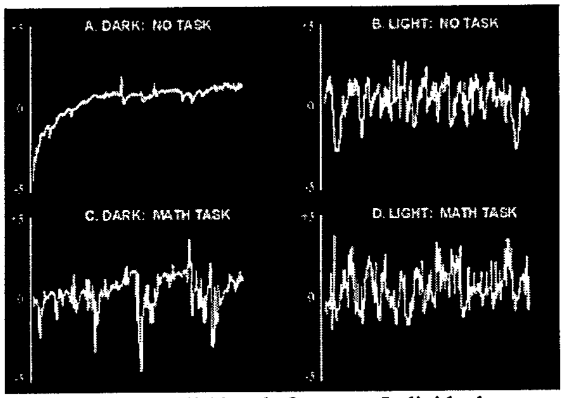


Figure 1. Pupil Signals from one Individual.

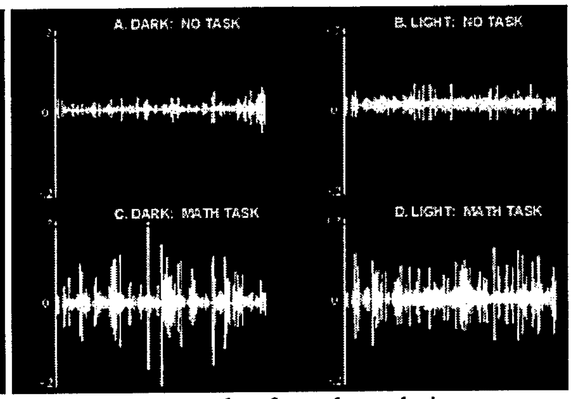


Figure 2. Results of wavelet analysis

As can be clearly seen in Figure 2, the wavelet procedure underlying the ICA filters out the light reflex, leaving only the desired dilation reflex that accompanies cognitive effort. The values plotted in the upper portion of the figure, i.e., light and dark conditions with no cognitive task, are essentially zero while those in the lower portion have many large non-zero spikes.

These four conditions have been replicated in several experiments across groups of participants, and all yielded essentially the same results: significant task difference (task versus no task) and non-significant light different (dark versus light). In all experiments, the interaction between task and light was not significant.

Moreover, as demonstrated in the figures above, the ICA can be computed for each individual. For example, in the experiment reported in the manuscript cited below, the main result of higher ICA for the task condition than the no task condition was observed in 22 of 23 participants. A simple binomial test shows that this outcome is extremely unlikely if the conditions are equivalent.

Simple Laboratory Experiments

A number of validating studies have been done, many based on reported laboratory experiments in the literature. The purpose of these studies was twofold: to determine whether similar overall results were found when compared to the original tests and to evaluate the size and location of the ICA across the various dimensions of the tasks. These tasks include: simple visual arithmetic problems; anagrams (with 3-8 letters); working memory tasks of digits, letters, shapes, and colors (sequences of 2-7 each); spatial reasoning (Raven Progressive Matrices), and visual search tasks. Results of several of these studies are reported in Marshall, Davis, & Knust (2003).

Equipment Comparison

The Index of Cognitive Activity was developed using one eye tracking system. Over the lifetime of this grant, the ICA was evaluated with several different systems. Specifically, data were collected using the Applied Science Laboratories 4000 Head Mounted System, the EyeLink I System supplied by SensoMotoric Instruments, Inc., and the EyeLink II System developed by SR International. The EyeLink System was originally developed by the SR group and then licensed for several years to SMI. SR International now exclusively manufactures and markets the EyeLink system. In addition, colleagues from other institutions have provided pupil recordings from other systems including an ISCAN RK406 pupillometer, and the ICA was successfully used for their data as well.

The sampling rates of the systems used at SDSU varied from 250 Hz (EyeLink II) to 60 Hz (ASL). With minor adjustments for the varying sampling rates, the Index of Cognitive Activity was successfully computed using data from all of these systems. It appears to be robust across sampling rates of 60 Hz and above.

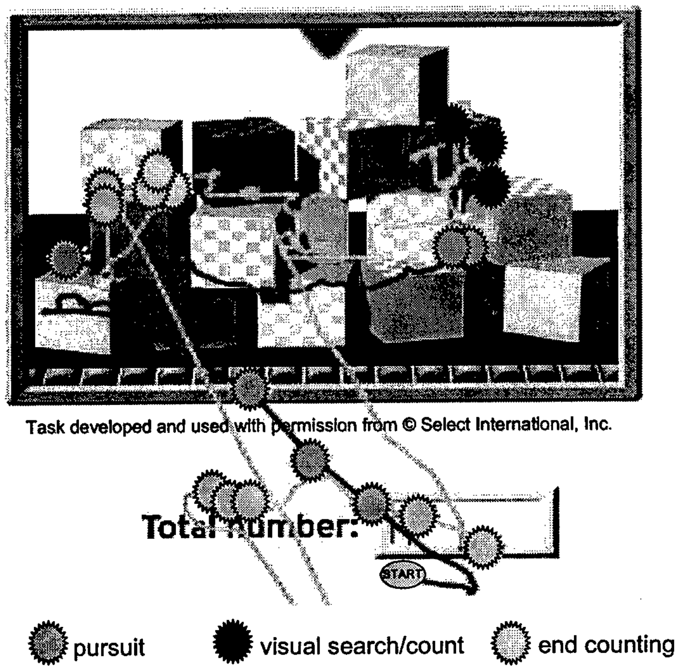
Applications of the Index of Cognitive Activity

The techniques developed here have been applied in several settings in which complex tasks are utilized. These tasks typically require the operator to maintain situation awareness and to respond to unusual events as they occur. These tasks often have immediate real-world counterparts. Two examples are described below.

<u>Screening/Visual Search</u>

For example, early in this program we tested extensively with a conveyor-belt task in which items were presented scrolling across the display from left to right. The operator's task was to search the display and determine the number of items, which varied in number, color, and shape complexity. Items could also block other items. An example is shown in the figure on the next page.

Eye Movements and Locations of Abrupt Pupillary Changes

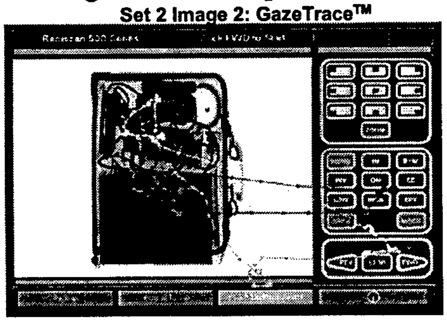


The Index of Cognitive Activity reflected the intensity of the task, rising with the number and difficulty of items. Moreover, it was evident from the eye movements during the task that the increased cognitive effort came not from the counting component itself but rather from the visual search to see every item and from the effort of holding and retrieving the final result in working memory.

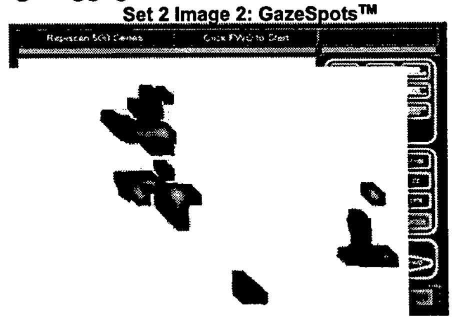
This relatively simple task has immediate transfer to the vital application of airport security screening. In baggage screening, items move across the display much as they did in our conveyor-belt task. Both tasks have the complexity added by shape, color, and overlapping presentation of items. Both require visual search and rapid decision. An example from an x-ray screening system is shown below. We are pursuing additional funding to continue this research.

Scanning Patterns of Operator Searching Luggage

Set 2 Image 2: GazeTrace™



Set 2 Image 2: GazeSpots™

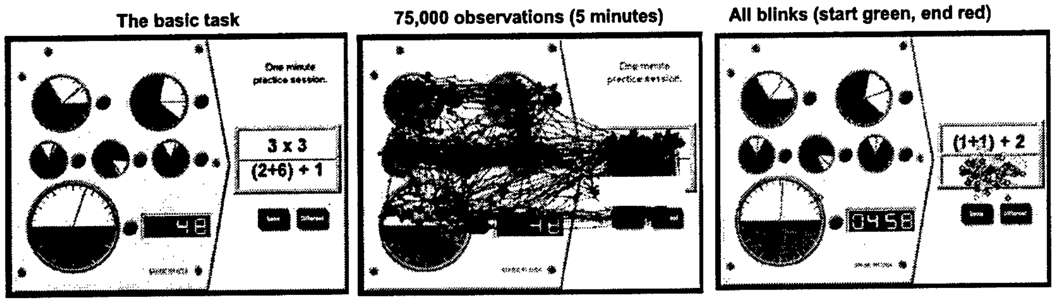


<u>Dual-Attention/Multiple Displays</u>

Similarly, we focused in the early periods of the program on a simple dual-attention task in which the operator was asked to monitor six gauges that were displayed on the left side of the display and simultaneously scan two arithmetic expressions presented on the right side of the display. His response to the left side was to adjust the needles within each gauge as required; his response to the right side was to compute each expression and indicate whether the totals were the same or different. The figure below shows the task on the left, all eye movements for one five-minute session in the middle, and the location of all blinks on the right. The Index of Cognitive Activity is sensitive to the difficulty of the comparison of the arithmetic expressions as well as to the speed with which the gauges changed.

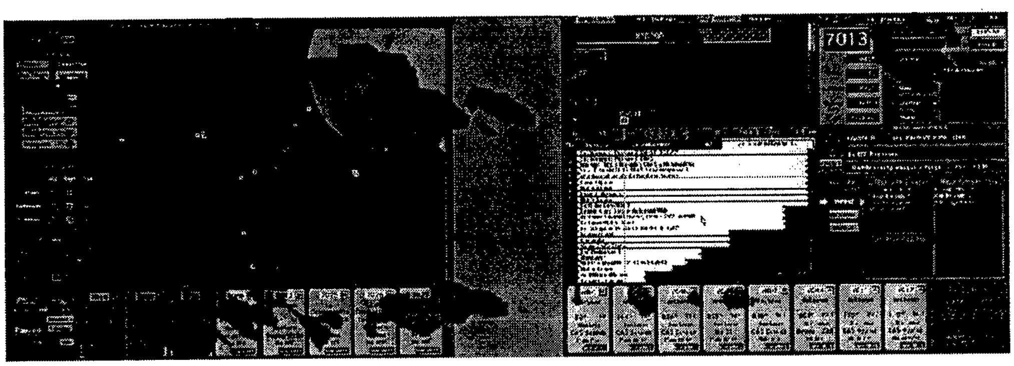
Eye Movements and Blinks during Dual Attention Task

The basic task | **75,000 observations (5 minutes)** | **All blinks (start green, end red)**



Task developed and used with permission from © Select International, Inc.

The use of multiple displays is common in applied military settings. For example, the figure below comes from the TADMUS Program of the Office of Naval Research.



Viewing pattern for one officer using the Decision Support System created for the TADMUS Program

The TADMUS display shows the aggregated viewing pattern of one officer working with the two displays for 30 minutes. The shaded areas indicate where attention was focused during the scenario. The Index of Cognitive Activity was successfully used with the DSS under funding from ONR to show that mental effort increases as expected under conditions of uncertainty.

Impact and Technology Transfer

As part of the overall DURIP project, we have created a new metric for measuring the amount of cognitive effort required by an operator in a variety of settings. The metric can be computed off-line following data collection or it can be computed in real-time as the data are being recorded. As is often the case with new measures, it has taken considerable time and effort to create the metric, validate it, and gain acceptance of its use.

One measure of its acceptance is the use of the Index of Cognitive Workload in several new projects. For example, it was the foundation on which two DARPA contracts were issued in 2002, one to San Diego State University for additional research about the properties of the index itself and a second to EyeTracking, Inc. for integration with EEG data. EyeTracking, Inc., was founded by Principal Investigator Sandra Marshall and others in her research lab in 1999, and it has become a well-known provider of eye-tracking services including usability studies, training assessments, and interface evaluations. Originally, EyeTracking, Inc. offered primarily eye-movement analysis. It now also makes available analysis of changes in cognitive effort as reflected by pupil diameter. Most recently, EyeTracking, Inc. has subcontracts with The Boeing Company and Lockheed Martin Advanced Technology Labs to continue the DARPA research efforts in real-time cognitive workload assessment.

Physiological Studies Associated with Cognitive Tasks

The major focus of the Neurokinetic laboratory's participation in the Argus project deals with physiological measurements associated with performance using the same Gopher Task that is being used at George Mason University. The research question deals with identifying various physiological factors that may or may not correlate with initial observations made by the George Mason team.

The George Mason team has quantitated a phenomenon in subjects participating in the Gopher Task. Specifically, the group identified a decrease in response time during the I+1 phase of the task. The SDSU group recorded eye blink, surface electromyograms from the extensor and flexor of the forearms, finger movement and finger force using the same task.

The data were as follows: Eye blinks occur more frequently during the Instruction and Instruction +1 phase. There are two interpretations for these data. First, the eye blink may occur during the instruction phase since the subject has to read a long passage and the I+1 phase also involves some reading. Second, the subjects may be involved in some process of cognition that is terminated by the eye blink.

The longer reaction times in the Instruction and I+1 phase may be due to a number of physiological or cognitive factors. Subjects had varying biomechical methods of responding with a number of them cocking their finger in anticipation. In addition, simultaneously, there was an associated eye blink in some subjects. It was tentatively concluded that Finger biomechanics (e.g. finger cocking) did not contribute significantly

to the phenomenon since the amount of finger cocking was not that great during these phases. It was noticed and subsequently measured that subjects during the I+1 phase generated more force than at earlier stages. This observation was the springboard for evaluating subject's responses to various stimuli.
Assuming that the eye blink is the termination of some form of cognition, it was determined that 500 msec are spent on determining appropriate responses and 400 msec are used in the execution of the motor command. These figures must be considered tentative since it was determined that the processing time of the computer influenced the time reported by the Gopher program.

The motion of the finger during depression has a low frequency of motion of approximately .05 hz and an associated high frequency of smaller amplitude of 12-15 Hz which is in the tremor frequency range . This high frequency may become exaggerated as the subjects become fatigued.

In addition, it was observed that the eyelid also had the tremor frequency of 12-15 hz.
The presence of the tremor in the eyelid suggests that neural mechanisms that control eyelid closure may play a roll in the genesis of eyelid tremor that could be exaggerated with fatigue.

Future research will examine the assumption underlying these data dealing with the stability of the various timed responses generated by the Gopher Task. since it was observed that the time varied between various computers. . To measure the stability of the timing reported by the Gopher task, photocells are being attached to the monitor and as soon as a choice is presented in the Gopher task, it will trigger automatically the depression of the C or M key. The research question that is being studied is whether the reaction time of this electrical response system stays constant. If the reaction time varies, this would indicate that there is an inherent time delay that varies in the Gopher program. The initial system has been built and is presently being tested.

In the next phase of the project, the research question that was asked dealt with the physiological changes that occur while the subject is undergoing a multitasking scenario. The scenario is a modification of the "gauges" task that Dr. Sandra Marshall is using. The scenario has two components: a set of gauges in which the task is to keep the needle in a certain area and a set of mathematical questions on the same screen asking the subject what is the correct answer. The mathematical component is on the right side of the screen and the gauges are on the left. The advantage of this task is that it takes only 5 minutes to complete and as such minimizes the criticism of fatigue effects associated with hour-long tasks.

Previous work by Dr. Marshall suggested that there were more eye blinks associated with the mathematical component of the gauges task, possibly suggesting that there is some form of cognition occurring followed by the eye blink. Another explanation could be that the subjects need to not blink during the gauges section of the task and blink only when they think they have more time. Alternatively, the blink rate could be due just to constant viewing on the crt screen. This present study modified the gauges task and asked what is

the timing associated with going from the gauges section to the math section. In addition, a tone was inserted into the program that began once the needle left the predetermined target area. The output from the new program produces a record of the time of when each button was depressed thus allowing for the determination of the path that the subject used in this scenario. The conditions that were explored dealt with the subject conducting the task with the tone on and off.

Subjects were seated and surface electrodes were placed on the forearm muscles, and back muscles. A specially designed mouse that measures finger force was used to measure the force of the finger as the subject pressed on the gauges button or the math answer. In addition, a video camera recorded the eye blink as the person performed the task.

To date, 15 subjects have participated in the study. The data suggest that there are various kinds of eye blinks associated with viewing the mathematical component. Sixty percent of the eye blinks are complete whereas the remainder are half blinks. The rationale for measuring eye blinks was to study whether or not they represented either a "cognitive punctuation" or a physiological function designed to keep the surface of the eye fluid. These studies did not allow for a clear elucidation of the genesis of the eye blink. The subjects performed significantly worse on both functions of the test when the tone was present. It was assumed that the tone would assist the subject in terms of notifying them when a needle was not in the target zone. Some subjects used the tone for that function, but most found it annoying. Force measurements of the fingers increased during the tone sequence.

EMG signals indicated that the use of timing of the forearm and depression of the key may not be a good indicator of response time since the EMGs of the back muscles (e.g. trapezius) were triggered sooner than the forearm muscles. Thus the CNS, when activated, sends the signal to certain predetermined muscle groups to respond to the stress and quantitating the response time as originating from the depression of the key may give misleading times in terms of this task.

Overall these studies suggest that during the gauges task, subjects blink more while on the mathematical section of the task but that the eye blink is not homogenous. The force measurements of the fingers on the mouse suggest that these measurements may be used as an indication of stress. As was noticed in previous studies, the generation of finger force may be an indication of stress such as uncertainty. Timing of the EMGs associated with the force production suggests that previous measurements of the reaction time may be in some ways misleading since the forearm signal is part of a predetermined muscle sequence.

The studies to date suggest that the gauges task may be stressful. However, to substantiate that point, measurement of heart rate will be conducted during the gauges task. Previous work using an electrocardiogram system that automatically measured "vagal" and "sympathetic" tone was inconclusive since the electrocardiac system required

minimal movement by the subject so as to minimize motion artifact. A new system was designed that has been incorporated into the recording system.

A new set of experiments will be conducted using the gauges system while measuring eye blink, heart rate, surface EMGs of the trapezius and forearm as well as finger force. The time of needle movement inside each gauge will be varied having a slow, intermediate and fast pace.

Continuing the studies on the stability of the timing of the Gopher task, it was concluded that varying computer configurations with different amounts of RAM significantly altered the time reported by the Gopher task. Comparing data from different computers using the same task needs to be approached cautiously due to this fact.

In the next phase of the project, the research question addressed dealt with the physiological changes that occur while the subject is undergoing a multitasking scenario. The scenario is a modification of the "gauges" task that Dr. Sandra Marshall is using. The scenario has two components: a set of gauges in which the task is to keep the needle in a certain area and a set of mathematical questions on the same screen asking the subject what is the correct answer. The mathematical component is on the right side of the screen and the gauges are on the left. The advantage of this task is that it takes only 5 minutes to complete and as such minimizes the criticism of fatigue effects associated with hour-long tasks.

Our work last period suggested that during the gauges task, subjects blink more while on the mathematical section of the task but that the eye blink is not homogenous. The force measurements of the fingers on the mouse suggest that these measurements may be used as an indication of stress. Timing of the EMGs associated with the force production suggests that previous measurements of the reaction time may be in some ways misleading since the forearm signal is part of a predetermined muscle sequence.

The initial observations concerning reaction time reported last period were confirmed. Before a person strikes a key in response to the question presented on the screen, the fingers are usually in a cocked position. This cocking occurs before the finger depresses the key suggesting that the CNS has poised the motor system in anticipation of the command. Reaction time has been classically used as a measure of the time the CNS processes data as well as the motor component. In our studies, the motor component is longer than those reported in the literature, thereby minimizing the time the CNS spends on computation. Additional work is continuing in this regard, this not all subjects demonstrate the cocking behavior. The amount of force used in these experiments varied with the amount of stress and/or fatigue of the subject.

These studies suggested that the gauges task may be stressful. In the past period, we have been engaged in substantiating that point through measurement of heart rate during the gauges task. Previous work using an electrocardiogram system that automatically measured "vagal" and "sympathetic" tone was inconclusive since the electrocardiac system required minimal movement by the subject. A new system was designed that has

been incorporated into the recording system. This equipment was then used to conduct a new set of experiments using the gauges system while measuring eye blink, heart rate, surface EMGs of the trapezius and forearm as well as finger force. The time of needle movement inside each gauge will be varied having a slow, intermediate and fast pace.

In this new work, subjects were outfitted with an electrocardiogram system as well as EMG recording system of their extensor and flexor muscles of the forearm. Analysis of these data recorded from subjects as they performed the gauges task showed an overall increase in both EKG and EMG amplitude that subsided as the subjects became more skilled. Subjects who returned for consecutive studies showed less of an increase on each succeeding visit. Although this was not thoroughly studied, it was noted that those individuals who had a history of playing computer games did not show any significant increase in physiological parameters studies.

While conducting experiments using the gauges task, we noticed that subjects had a certain motor pattern as they approached the buttons to reset the gauges.
High speed video cameras recorded the motion of the subjects as they moved the mouse/joystick during the gauges task. Specifically, subjects demonstrated a major deceleration followed by two smaller decelerations as they approached the reset button. The smaller decelerations occurred immediately before target acquisition with very short times (15 msec) indicating that these decelerations were programmed before the start of the movement. The fastest reflex recorded in the body is the stretch reflex that has a time of 30-50 msec. Thus, for the CNS to correct for target acquisition, it has to make those decisions in a preprogrammed mode. These small decelerations may be a reflection of speed /accuracy tradeoffs. These studies are being replicated and, if duplicated, could shed light on the motor component mechanisms involved in target acquisition.

The final period of the project witnessed some significant advances in our understanding of the mechanisms behind motor movement of the finger as it depresses a key. Previous studies conducted with the George Mason group suggested that finger movement during various psychological paradigms measuring key depression was not a simple up and down movement suggesting some preprocessing of the motor command previous to motor execution.

Studies were conducted to elucidate more about this mechanism(s) and the findings are as follows:

1. Before finger impact on the key, the finger undergoes a deceleration immediately before impact. This finding corresponds with what we have reported with large movements of the arm or leg. Various experimental arrangements were used to document this finding using high speed cameras recording the data at 500-1000 frames/sec as well as surface electromyograms. Studies consisted of

    a. Persons hitting only one key (n=12)

    b. Persons hitting sequential keys(n=12)

c. Persons hitting a suspended small object that the person had hit with their finger(n=6)

2. The timing of the deceleration is 10-20 msec which is too fast for any form of feedback mechanism.

3. Upon impact, the finger position is readjusted. Initially, there is a small depression and elevation of the finger followed by a large depression and elevation of the finger. The interpretation of these data suggests that the finger is adjusting for hitting the key by an initial adjustment. The timing of these two movements is such at they cannot be explained on the basis of feedback loops from the skin or visual systems.

4. Biomechanical Data: The above studies were subsequently followed by having a joint video and force measurement studies. Force was measured on a keyboard in which force transducers had been placed on the key. Signals from the transducers were sampled at 1000 Hz. The force measurements supported the video data in that there was an initial small force profile followed by a large force profile. These data support the kinematic data indicating that the finger hits the key twice. (N=10)

Significance of the data:

The above studies suggest that the execution of finger movements are preprogrammed and that there is minimal to no adjustments occurring at the key strike. Thus, before impact, the CNS has made the necessary adjustments to hit the target. These observations led to the next set of studies that dealt with the force development of the fingers to the point of fatigue to asses whether the phenomenon (e.g. two peaks in force profile) mentioned above were modified by fatigue.

Twelve subjects were asked to sequentially generate force on a keyboard that had five force transducers applied to each key. Subjects were asked to depress the keys as a specific rate determined by a metronome of 3/sec until they were fatigued.

Data indicated the following:

1. Finger fatigue occurred within 3-5 minutes and the kinematic and force records indicate that the initial movements to readjust the finger as listed above become exacerbated.

2. Thumb fatigue the first followed by ring and little finger.

Kinematic analyses of single finger movement-three dimensional analyses suggest the following. First, during the above studies, it became apparent that the finger had a complex pattern of movement. Two high speed cameras recorded the movement of the finger as it hit the key.

Second, analyses indicated that the finger had a parabolic movement pathway. The finger descended in a vertical line and after hitting the key it would return to the initial elevated position by a circular path. In addition as it descended, the finger would have major movement in the Z plane indicated an increased flexed movement of the finger inward. These measurements became enhanced with fatigue.

Significance of overall research activity:

Many psychophysiological studies dealing with elucidating various questions dealing with cognition or brain processing have assumed that the timing from the presentation of stimuli to the execution of some motor command had distinct components. Our studies suggest that at least on the motor side that when a command is given to hit the keyboard, there are adjustments made to strike the key that are preprogrammed. This preprogramming takes into account the target and executes the movement with a subcomponent that establishes contact with the target. Thus the finger will have a decrease in velocity or acceleration previous to hitting the target. In addition, there may be some minor increases and decrease in movement before the major deceleration. This phenomenon that we have quantitated with the finger movements we have also seen in larger faster movements such as the baseball swing and the soccer kick. Overall these data suggest that a common motor response in many cases is preprogrammed and that reaction time measurements must be considered in this light.

Just as important, the measurement of finger force suggests that studies that propose to model cognitive processing time should consider this variable. The striking of a key is not sufficient data to model cognitive processing since the major function of the motor side also include force, acceleration, velocity and displacement measurements. Monitoring only the time that it takes to strike a key after some stimuli is similar to monitoring whether a boxer just strikes his opponent. Without force measurements, most models will not consider the more important aspect of cognitive decision making.

These data suggest that in some experiments, the use of motor movement as a sign of central nervous system timing may actually have occurred many milliseconds previous to the motor movement and that the motor system is executing a motor sequence that cannot be altered by sensory signals. Thus the use of reaction time measurements may be misleading in terms of understanding motor times. In addition, experiments that continue for an hour introduce fatigue and the data may be modified due to the effects of fatigue, and/or boredom.

Personnel Involved in this Effort

Leonard Adelman
Erik Altmann
B. Barnes
Deborah Boehm-Davis
M. Camarena
Ronald Chong
J. M. Conte
Katie Cunningham
Cassie Davis
B. T. Dickson
Wai-tat Fu
Robert Gambill
Wayne Gray
DeVere Henderson
K. Ishizakas
Sandra Marshall
Sheryl Miller
Christopher Myers
P. Osborne
C. W. Pleydell-Pearce
Robert Pozos
Susan Schnipke
Michael Schoelles
Ryan Snead
Mark Todd
Cedric Yeo

Papers

Adelman, L. & Gambill, R. (in press). Examining the effect of reward structure and system interface on operators' information-processing strategies under time pressure. *Proceedings of the 20003 IEEE Systems, Man, and Cybernetics Conference*, Washington, DC, October 2003.

Adelman, L., Henderson, D., & Miller, S. (2001). Vicarious functioning in teams. In K.R. Hammond and T.R. Stewart (Eds.), *The essential Brunswik: Beginnings, explications, and applications* (pp. 416-423). NY: Oxford University Press.

Adelman, L., Henderson, E.D., and Miller, S.L. (1998a). Examining the effects of workload and task complexity on team communication processes and decision making (ARCH Lab Report 98-02/10). Fairfax, VA: .George Mason University.

Adelman, L., Miller, S.L., Henderson, D., & Schoelles, M. (2003). Using Brunswikian theory and the Team Argus simulated task environment to study how hierarchical teams adapt to increasing levels of time pressure, *Acta Psychologica,* vol. 112, no. 2, 181-206.

Adelman, L., Miller, S.L., & Yeo, C. (accepted, 2003). Testing the effectiveness of icons for supporting distributed team decision making under time pressure. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans.*

Adelman, L., Miller, S.L., & Yeo, C. (2001). Testing the effectiveness of perceptualinterface components for supporting distributed team decision making. Seventeenth Annual Meeting of the International Brunswik Society, Orlando, FL, 15-16 Nov. 2001.

Adelman, L., Yeo, C., & Miller, S.L. (in preparation). Using Brunsikian theory, the lens model equation, and path models to understand the effects of computer displays and time pressure on the performance of distributed teams. In A. Kirlick, Ed., *Working with Technology in Mind: Brunswikian Resources for Cognitive Science and Engineering*, NY: Oxford University Press.

Adelman, L., Yeo, C., & Miller, S.L. (2000). Relative effectiveness of different interfaces to ameliorate the negative effects of time pressure on team performance. Sixteenth Annual Meeting of the International Brunswik Society, Berlin, Germany, 20-22 July 2000.

Altmann, E. M. (2002). Functional decay of memory for tasks. *Psychological Research-Psychologische Forschung, 66*(4), 287-297.

Altmann, E. M. (2001). Near-term memory in programming: a simulation-based analysis. *International Journal of Human-Computer Studies, 54*(2), 189-210.

Altmann, E. M. (2000). Memory in chains: Modeling primacy and recency effects in memory for order. In *Proceedings of the Twenty-Second Annual Meeting of the Cognitive Science Society*, Mahwah, NJ: Lawrence Erlbaum Press, 31-36.

Altmann, E. M. (2000) Memory in chains: A dual-code associative model of positional uncertainty. In N. Taatgen & J. Aasman (Eds.), *Proceedings of the Third International Conference on Cognitive Modeling* (pp. 9-16). Veenendal, NL: Universal Press.

Altmann, E. M. (1999). *Paying attention by preparing to forget.* (Unpublished abstract).

Altmann, E. M. (1999). Near-term memory in programming: A simulation-based analysis. (Manuscript submitted for publication).

Altmann, E. M., Cleeremans, A., Schunn, C. D., & Gray, W. D. (Eds.). (2001). *Fourth International Conference on Cognitive Modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.

Altmann, E. M., & Davidson, D. J. (2001). An integrative approach to Stroop: Combining a language model and a unified cognitive theory. In J. D. Moore & K. Stenning (Eds.), *Twenty-Third Annual Conference of the Cognitive Science Society* (pp. 21-26). Hillsdale, NJ: Lawrence Erlbaum Associates.

Altmann, E. M., & Gray, W. D. (2002). Forgetting to remember: The functional relationship of decay and interference. *Psychological Science, 13*(1), 27-33.

Altmann, E. M., & Gray, W. D. (2000). Managing attention by preparing to forget, *Human Factors and Ergonomics Society 44th Annual Meeting* (pp. 152-155). Santa Monica, CA: Human Factors and Ergonomics Society.

Altmann, E. M., and Gray, W. D. (2000) An integrated model of set shifting and maintenance. In N. Taatgen & J. Aasman (Eds.), *Proceedings of the Third International Conference on Cognitive Modeling* (pp. 17-24). Veenendal, NL: Universal Press.

Altmann, E. M., & Gray, W. D. (1999). Functional decay in serial attention, *Sixth ACT-R Workshop*. Fairfax, VA: ARCH Lab.

Altmann, E. M., & Gray, W. D. (1999). Serial attention as strategic memory. In M. Hahn & S. C. Stoness (Eds.), *Twenty-First Annual Conference of the Cognitive Science Society* (pp. 25-30). Hillsdale, NJ: Lawrence Erlbaum Associates.

Altmann, E. M., & Gray, W. D. (1998). Pervasive episodic memory: Evidence from a control-of-attention paradigm. In Proceedings of the Twentieth Annual Conference of the Cognitive Science Society (pp. 42-47). Hillsdale, NJ: Erlbaum.

Altmann, E. M., Gray, W. D., Lipps, A., & Trickett, S. (1998). Individual differences in the strategic processing of brief interruptions (ARCH Lab Report 98-03/04). Fairfax, VA: George Mason University.

Altmann, E. M., & John, B. E. (1999). Episodic indexing: A model of memory for attention events. *Cognitive Science, 23*(2), 117-156.

Altmann, E. M., & Schunn, C. D. (2002). *Integrating decay and interference: A new look at an old interaction.* Paper presented at the Twenty-fourth Annual Conference of the Cognitive Science Society, Mahwah, NJ.

Altmann, E. M. & Trafton, J. G. (1999). Memory for goals: An architectural perspective. In *Proceedings of the twenty first annual meeting of the Cognitive Science Society.*

Boehm-Davis, D. A., Gray, W. D., and Schoelles, M. The eye blink as a physiological indicator of cognitive workload. In *Proceedings of the International Ergonomics Association/Human Factors and Ergonomics Society 2000 Congress*, Santa Monica, CA: Human Factors and Ergonomics Society.

Fu, W.-T. (2003). Adaptive planning in problem solving - cost-benefit tradeoffs in bounded rationality. Unpublished doctoral dissertation, George Mason University, Fairfax, VA.

Fu, W.-t. (2001). ACT-PRO: Action protocol tracer -- a tool for analyzing discrete action protocols. Behavior Research Methods, Instruments, & Computers, 33(2), 149-158.

Fu, W.-t., & Gray, W. D. (2003). Why do inefficient procedures persist in interactive tasks? Regularities and frequencies in display-based operator schemata. *Manuscript submitted for publication.*

Fu, W.-t., & Gray, W. D. (2000). Memory versus Perceptual-Motor Tradeoffs in a Blocks World Task. In L. R. Gleitman & A. K. Joshi (Eds.), Twenty-second Annual Conference of the Cognitive Science Society (pp. 154-159). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fu, W.-t., & Gray, W. D. (1999). ACT-PRO: Action protocol tracer -- a tool for analyzing simple, rule-based tasks, Sixth ACT-R Workshop. Fairfax, VA: ARCH Lab.

Fu, W.-t., & Gray, W. D. (1999). Redirecting direct-manipulation, or, what happens when the goal is in front of you but the interface says to turn left?, CHI 99 Extended Abstracts (pp. 226-227). New York: ACM Press.

Gray, W. D. (2003). Soft Constraints on Interactive Behavior: Perspectives, Recent Research, and Future Directions. Paper presented at the IK2003, Guhe, GE.

Gray, W. D. (2002, May 22-24). Global and local adaptations to soft constraints on interactive behavior. Paper presented at the ONR Workshop on Attention, Perception and Data Visualization, Fairfax, VA.

Gray, W. D. (2002). Simulated task environments: The role of high-fidelity simulations, scaled worlds, synthetic environments, and microworlds in basic and applied cognitive research. Cognitive Science Quarterly, 2(2), 205-227.

Gray, W. D. (2002). Soft Constraints on Interactive Behavior: Perspectives, Recent Research, and Future Directions, Invited talk given to the Cognitive Science Department of Rensselaer Polytechnic Institute. Troy, NY.

Gray, W. D. (2000, May). Following the path of cognitive least-effort may be a lot of work: Is cognition a local optimizer? Paper presented at the Mid-Western Psychological Society, Chicago, IL.

Gray, W. D. (2000). The nature and processing of errors in interactive behavior. Cognitive Science, 24(2), 205-248.

Gray, W. D. (1999). Scaled worlds: Tractable, realistic, and engaging? Paper presented at the Scaled Worlds '99: Inaugural International Synthetic Task Development Conference, Athens, GA.

Gray, W. D., & Altmann, E. M. (2001). Cognitive modeling and human-computer interaction. In W. Karwowski (Ed.), International encyclopedia of ergonomics and human factors (Vol. 1, pp. 387-391). New York: Taylor & Francis, Ltd.

Gray, W. D., & Boehm-Davis, D. A. (2000). Cognitive analysis of dynamic performance: Cognitive process analysis and modeling. Paper presented at the Human Factors and Ergonomics Society, San Diego, CA.

Gray, W. D., & Boehm-Davis, D. A. (2000). Milliseconds Matter: An introduction to microstrategies and to their use in describing and predicting interactive behavior. Journal of Experimental Psychology: Applied, 6(4), 322-335.

Gray, W. D., & Boehm-Davis, D. A. (1998). Argus Prime: Dynamic micro-strategies, concurrent task execution, and cognitive workload (No. ARCH Lab Report 98-05/18). Fairfax, VA: George Mason University.

Gray, W. D., & Boehm-Davis, D. A. (1997). Cognitive analysis of dynamic performance: Cognitive process analysis and modeling. Paper presented at the Human Factors and Ergonomics Society, Albuquerque, NM.

Gray, W. D., & Boehm-Davis, D. A. (1996). Workshop on Cognitive Task Analysis: The GOMS family. Paper presented at the Human Factors and Ergonomics Society, Philadelphia.

Gray, W. D., & Fu, W.-t. (2003). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. Manuscript submitted for publication.

Gray, W. D., & Fu, W.-t. (2001). Ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head: Implications of rational analysis for interface design. Paper presented at the ACM CHI'01 Conference on Human Factors in Computing Systems, New York.

Gray, W. D., & Fu, W.-t. (2001). Ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head: Implications of rational analysis for interface design. CHI Letters, 3(1), 112-119.

Gray, W. D., & Fu, W.-t. (2000). The influence of source and cost of information access on correct and errorful interactive behavior. In L. R. Gleitman & A. K. Joshi (Eds.), Twenty-second Annual Conference of the Cognitive Science Society (pp. 663-668). Mahwah, NJ: Lawrence Erlbaum Associates.

Gray, W. D., & Fu, W.-t. (2000, November). Memory versus Perceptual-Motor Tradeoffs in a Blocks World Task. Paper presented at the Conference of the Psychonomics Society, New Orleans.

Gray, W. D., Palanque, P., & Paternò, F. (1999). Introduction to the Special Issue on: Interface Issues and Designs for Safety-critical Interactive Systems. ACM Transactions on Computer-Human Interaction, 6(4), 309-310.

Gray, W. D., & Salzman, M. C. (1998). Damaged merchandise? A review of experiments that compare usability evaluation methods. Human-Computer Interaction, 13(3), 203-261.

Gray, W. D., & Salzman, M. C. (1998). Repairing damaged merchandise: A rejoinder. Human-Computer Interaction, 13(3), 325-335.

Gray, W. D., & Schoelles, M. J. (2003). The nature and timing of interruptions in a complex, cognitive task: Empirical data and computational cognitive models. In R. Alterman & D. Kirsch (Eds.), 25th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society.

Gray, W. D., & Schoelles, M. J. (2000, November). Argus: A suite of tool for research in complex cognition. Paper presented at the Society for Computers in Psychology, New Orleans.

Gray, W. D., Schoelles, M. J., & Fu, W.-t. (2000). Modeling a continuous dynamic task. In N. Taatgen & J. Aasman (Eds.), Third International Conference on Cognitive Modeling (pp. 158-168). Veenendal, The Netherlands: Universal Press.

Gray, W. D., Schoelles, M. J., & Fu, W.-t. (1999). Modeling microstrategies in a

continuous dynamic task. Manuscript submitted for publication IJHCS.

Gray, W. D., Schoelles, M. J., & Myers, C. W. (2002). Computational cognitive models ISO ecologically optimal strategies, 46th Annual Conference of the Human Factors & Ergonomics Society (pp. 492-496). Santa Monica, CA: Human Factors & Ergonomics Society.

Gray, W. D., & Schunn, C. D. (Eds.). (2002). Proceedings of the 24th Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Erlbaum.

Gray, W. D., Young, R. M., & Kirschenbaum, S. S. (1997). Introduction to this Special Issue on Cognitive Architectures and Human-Computer Interaction. Human-Computer Interaction, 12(4), 301-309.

Henderson, E.D. (1999). Examining the Adaptive Decision Making Behavior of Distributed Hierarchical Teams Under High Temporal Workload. (George Mason University dissertation; anticipated completion by May 1999.)

Henderson, E.D. (1999). *Model for Adaptive Decision Making Behavior of Distributed Hierarchical Teams Under High Temporal Workload.* (Ph.D. dissertation) Fairfax, VA: George Mason University.

Ishizaka, K., Marshall, S. P., & Conte, J. M. (2001). Individual differences in attentional strategies in multitasking situations. *Human Performance.*

John, B. E. & Altmann, E. M. (1999). The power and constraint provided by an integrative cognitive architecture. In *Proceedings of the second international conference on cognitive science and the 16th annual meeting of the Japanese cognitive science society joint conference.*

Marshall, S. P. (2002). The Index of Cognitive Activity: Measuring cognitive workload. In *Proceedings of the 2002 IEEE 7th Conference on Human Factors and Power Plants* (pp. 7.5 – 7.9). New York: IEEE.

Marshall, S. P. Cognitive workload and pupil dilation: A useful and usable indicator of cognitive activity. In E. Granholm & S. Steinhauer (chairs), *In the mind's eye: A view through the pupil at cognition in psychopathology.* Symposium presented at the 40th Annual Meeting of the Society for Psychophysiological Research, San Diego, October 2000.

Marshall, S. P., Davis, C. L., Knust, S. R. (2003). *The Index of Cognitive Activity: Estimating Cognitive Effort from Pupil Dilation.* (manuscript submitted for publication; currently under revision).

Marshall, S. P., Pleydell-Pearce, C. W., & Dickson, B. T. (2003). Integrating psychophysiological measures of cognitive workload and eye movements to detect

strategy shifts. In *Proceedings of the Thirth-Sixth Annual Hawaii International Conference on System Sciences.* Los Alamitos, CA: IEEE.

Miller, S.L., Adelman, L., Henderson, E.D., Schoelles, M., and Yeo, C. (2000). Team decision-making strategies: Implications for designing the interface in complex tasks. In *Proceedings of the International Ergonomics Association/Human Factors and Ergonomics Society 2000 Congress*, Santa Monica, CA: Human Factors and Ergonomics Society.

Miller, S.L., Adelman, L., Yeo, C., & Henderson, E.D. (2001). The effect of time pressure on speed and accuracy in team decision making: Designing the interface to support team strategies. To be submitted to *Human Factors* for review in September.

Pozos, R. S. (1999). *Psychophysiological Studies*. SDSU Working Paper, San Diego, CA.

Schnipke, S. K., Altmann, E. M., & Gray, W. D. (1998). Factors in the adoption and maintenance of suboptimal micro-strategies (ARCH Lab Report 98-03/05). Fairfax, Va: George Mason University.

Schoelles, M. J. (2002). Simulating Human Users in Dynamic Environments. Unpublished doctoral dissertation, George Mason University, Fairfax, VA.

Schoelles, M. J., & Gray, W. D. (2003, July). Cognitive constraints in dual-task performance: Implications of the ACT-R architecture. Paper presented at the 10th Annual ACT-R Workshop, Carnegie Mellon University, Pittsburgh, PA.

Schoelles, M. J., & Gray, W. D. (2003). Top-down versus bottom-up control of cognition in a task switching paradigm, Fifth International Conference on Cognitive Modeling. Bamberg, GE.

Schoelles, M. J., & Gray, W. D. (2002). Argus Prime: Adventures with ACT-R 5.0. Retrieved March 15, 2002, from http://www.dtic.mil/AFRL/afrl.html

Schoelles, M. J., & Gray, W. D. (2001). Argus: A suite of tools for research in complex cognition. Behavior Research Methods, Instruments, & Computers, 33(2), 130-140.

Schoelles, M. J., & Gray, W. D. (2001). Decomposing interactive behavior. In J. D. Moore & K. Stenning (Eds.), Twenty-Third Annual Conference of the Cognitive Science Society (pp. 898-903). Mahwah, NJ: Lawrence Erlbaum Associates.

Schoelles, M. J., & Gray, W. D. (2000). Argus Prime: Empirical test of the Argus Prime ACT-R/PM model at the unit task level, Seventh ACT-R Workshop. Pittsburgh, PA: Carnegie Mellon University.

Schoelles, M. J., & Gray, W. D. (2000). Argus Prime: Model-driven progressive deepening in search of the origins of cognitive workload. Manuscript submitted for

publication.

Schoelles, M. J., & Gray, W. D. (2000). Argus Prime: Modeling emergent microstrategies in a complex simulated task environment. In N. Taatgen & J. Aasman (Eds.), Third International Conference on Cognitive Modeling (pp. 260-270). Veenendal, NL: Universal Press.

Schoelles, M. J., & Gray, W. D. (2000). Modeling embodied cognition in a complex, real-time task. In L. R. Gleitman & A. K. Joshi (Eds.), Poster Twenty-second Annual Conference of the Cognitive Science Society (pp. 1052). Mahwah, NJ: Lawrence Erlbaum Associates.

Schoelles, M. J., & Gray, W. D. (1999). Modeling an interactive dynamic task: Argus Prime, Sixth ACT-R Workshop. Fairfax, VA: ARCH Lab.

Schoelles, M. J., & Gray, W. D. (1998). Argus, A System for Varying Cognitive Workload. Paper presented at the Human Factors and Ergonomics Society 42th Annual Meeting, Chicago.

Trafton, J. G., Marshall, S., Mintz, F., & Trickett, S. B. (2002). Extracting explicit and implict information from complex visualizations. In M. Hegarty, B. Meyer, & H. Narayanan (Eds.), Diagramatic representation and inference (p. 206-220). Berlin Heidelberg: Springer-Verlag.

Presentations

Adelman, L., Henderson, D., & Miller, S. (1999, April). Examining How Teams Adapt to Increasing Time Pressure. Presented at Georgetown University, Washington, DC.

Adelman, L., Henderson, D., & Miller, S. (1998, November). Vicarious functioning in teams. Presented at the Fourteenth Annual Meeting of the Brunswick Society.

Adelman, L., Miller, S., Henderson, D., Schoelles, M., & Foster, D. (1999, June??). "Synthetic worlds": Using the Argus system to help understand how hierarchical teams adapt to increasing time pressure.

Adelman, L., Yeo, C., and Miller, S.L. (2000). Relative effectiveness of different interfaces to ameliorate the negative effects of time pressure on team performance. Presented at the 16th Annual Meeting of the International Brunswik Society, Berlin, Germany, 20-22 July 2000.

Altmann, E. M. (2000). *Memory in chains: Modeling primacy and recency effects in memory for order*. Paper presented at the Twenty-Second Annual Meeting of the Cognitive Science Society, Philadelphia, PA.

Altmann, E. M. (2000) *Memory in chains: A dual-code associative model of positional uncertainty*. Paper presented at the Third International Conference on Cognitive Modeling, Groningen, NL.

Altmann, E. M. (1999, March). Paying attention by preparing to forget. ??

Altmann, E. M. (??). Near-term memory in programming: A simulation-based analysis.

Altmann, E. M., and Gray, W. D. (2000) *An integrated model of set shifting and maintenance*. Paper presented at the Third International Conference on Cognitive Modeling, Groningen, NL.

Altmann, E. M., and Gray, W. D. (2000) *Managing attention by preparing to forget.* Paper presented at the International Ergonomics Association/Human Factors and Ergonomics Society 2000 Congress, San Diego, CA.

Altmann, E. M. & Gray, W. D. (1999, June). Managing attention by preparing to forget. Presented at the ONR Attention-Management Workshop, Arlington, VA.

Altmann, E. M. & Gray, W. D. (1999). Serial Attention as Strategic Memory In *Proceedings of the twenty first annual meeting of the Cognitive Science Society.*

Altmann, E. M. & Gray, W. D. (1999, August). Functional decay in serial attention. Presented at the ACT-R Workshop, George Mason University, Fairfax, VA.

Altmann, E. M. & Trafton, J. G. (1999). Memory for goals: An architectural perspective. Presented at the Twenty-first Annual Meeting of the Cognitive Science Society.

Altmann, E. M. & Trafton, J. G. (1999, August). Functional encoding in memory for goals. Presented at the ACT-R Workshop. George Mason University, Fairfax, VA.

Altmann, E. M. & Trafton, J. G. (submitted to ??). Memory for goals in means-ends behavior.

Altmann, E. M. & Trafton, J. G. (1999). *Memory for goals: An architectural perspective.* Presented at the Twenty-first Annual Meeting of the Cognitive Science Society.

Altmann, E. M. & Trafton, J. G. (1999, August). *Functional encoding in memory for goals.* Presented at the ACT-R Workshop. George Mason University, Fairfax, VA.

Boehm-Davis, D. A. (2001, March 28). Testimony presented to Congress on funding for behavioral and social sciences research. Presented on behalf of the American Psychological Association..

Boehm-Davis, D. A. (1999, August). Heard a good (engineering psychology) story lately? Tell Congress. Paper presented at the Annual Meeting of the American Psychological Association, Boston, MA.

Boehm-Davis, D. A., Gray, W. D., and Schoelles, M.(2000) *The eye blink as a physiological indicator of cognitive workload.* Paper presented at the International Ergonomics Association/Human Factors and Ergonomics Society 2000 Congress, San Diego, CA.

Butler, K., Gray, W. D., & Jacob, R. J. K. (2001). Human-computer interaction: Introduction and overview, Tutorial presented at the ACM CHI '2001 Conference on Human Factors in Computing Systems.

Gray, W. D. (2000, May). *Following the path of cognitive least-effort may be a lot of work: Is cognition a local optimizer?* Paper presented at the Mid-Western Psychological Society, Chicago, IL.

Gray, W. D. (1999). Scaled Worlds: Tractable, Realistic, and Engaging? Paper presented at the Scaled Worlds '99: Inaugural International Synthetic Task Development Conference, Athens, GA.

Gray, W. D., & Altmann, E. M. (1998). Control of Attention as a Strategic Response to Limits of the Human Cognitive Architecture: Implications for Understanding Cognitive Workload. Invited talk presented at the Naval Research Laboratory, Nov. Washington, DC.

Gray, W. D. & Boehm-Davis, D. A. (April 19, 2001). Why Milliseconds Matter: Building models at the 0.100 sec (100 msec) level to explain human performance at the 10,000 sec (15 min) level. Presented at the Human Factors and Ergonomics Society Potomac Chapter Meeting, Arlington, VA.

Gray, W. D., & Boehm-Davis, D. A. (2000). *Cognitive analysis of dynamic performance: Cognitive process analysis and modeling.* Paper presented at the Human Factors and Ergonomics Society, San Diego, CA.

Gray, W. D., Boehm-Davis, D. A., John, B.E., and Kieras, D. (1999, May). *Cognitive modeling of dynamic tasks.* Workshop presented at Carnegie-Mellon University (concurrent with the Computer-Human Interaction Conference), Pittsburgh, PA.

Gray, W. D., & Fu, W.-t. (2001). Ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head: Implications of rational analysis for interface design. Paper presented at the ACM CHI'01 Conference on Human Factors in Computing Systems, New York.

Gray, W. D., & Fu, W.-t. (2000). *The influence of source and cost of information access on correct and errorful interactive behavior*, Paper presented at the Twenty-second Annual Conference of the Cognitive Science Society, Philadelphia, PA

Gray, W. D., & Schoelles, M. J. (2000, November). Argus: A suite of tool for research in complex cognition. Paper presented at the *Society for Computers in Psychology*, New Orleans.

Gray, W. D., & Schoelles, M. J. (November 2000). Argus – A flexible, multiple-cue probability task for modeling and studying individual and team cognitive workload, strategies, and microstrategies. Paper presented at the annual meeting of the *Society for Computers in Psychology*. New Orleans.

Gray, W. D., Schoelles, M. J., & Fu, W.-t. (2000). *Modeling a continuous dynamic task.* In N. Taatgen & J. Aasman (Eds.), Paper presented at the Third International Conference on Cognitive Modeling, Groningen, NL.

Gray, W. D., Schoelles, M., & Fu, W. (July 1998). *Milliseconds Matter*. Paper presented at the Fifth Annual ACT-R Workshop. Pittsburgh, PA: Carnegie Mellon University.

Marshall, Sandra. (2003, November). *New Techniques for Evaluating Innovative Interfaces with Eye Tracking.* Keynote address to be presented at the joint meeting of the 16th ACM Symposium on User Interface Software and Technology) and ICMI-PUI (the International Conference on Multimodal Interfaces), Vancouver, British Columbia.

Marshall, S. P. (2003, September). Measures of Attention and Cognitive Effort in Tactical Decision Making. To be presented at the Conference on Human Factors of Decision Making in Complex Systems, Dunblane, Scotland.

Marshall, S. P. (2002, September). The Index of Cognitive Activity: Measuring Cognitive Workload.. Presented at the 2002 IEEE 7th Conference on Human Factors and Power Plants. Scottsdale AZ.

Marshall, Sandra. (2002, May). *Cognitive and Instructional Applications for New Eye-Tracking Technologies.* Invited address presented at the International Conference on Application of Neuroscience Technology to Educational and Social Research, Hong Kong.

Marshall, S. (2001, July). *Eye Tracking: A Rich Source of Information for User Modeling*. Invited address presented at the 8th International Conference on User Modeling, UM2001, Sonthofen, Germany.

Marshall, S. P. (2000, April). *EyeTracking: A Tool for Tomorrow's Web-Based Training Designer.* Invited address presented at the WBT Producer Conference & Expo, San Diego.

Marshall, S. P. (2000, October). Using eye tracking to understand situational awareness. In H. Gigley (chair), Cognitive Aspects of Situation Awareness. Symposium presented at the Conference on Human Performance, Situation Awareness and Automation: User-Centered Design for the New Millennium. Savannah, GA

Marshall, S. P. (2000, October). Cognitive workload and pupil dilation: A useful and usable indicator of cognitive activity. In E. Granholm & S. Steinhauer (chairs), In the mind's eye: A view through the pupil at cognition in psychopathology. Symposium presented at the 40th Annual Meeting of the Society for Psychophysiological Research, San Diego..

Marshall, S. P. (1999, July 21). *Cognitive Applications of New Computational Technologies in Eye Tracking*. Invited address presented to the International Artificial Intelligence in Education Society at the AI-ED 99 Conference in Le Mans, France.

Marshall, S. P., & Davis, C. L. (1999, September). Monitoring Eye Movements in Tactical Decision Making. Poster presented at the European Eye Movement Conference, Utrecht, Netherlands.

Marshall, S.P., Knust, S. R., & Ishizaka, K. (2000, July). Using eye activity to study the cognitive processes of individual differences. Presented at the Annual Meeting of the International Ergonomics Association and Human Factors and Ergonomics Society, San Diego.

Marshall, S. P., Pleydell-Pearce, C., & Dickson, B. T. (2003, January). Integrating psychophysiological measures of cognitive workload and eye movements to detect strategy shifts. Presented at the 36th Annual Hawaii International Conference on System Sciences (HICSS). Kona, Hawaii.

Marshall, S. P., & St. John, M. (2003, January). The Index of Cognitive Activity as a Measure of Cognitive Workload on the Warship Commander Task (WCT). Poster presented at the joint session of DARPA's Bio Bionics Augmented Cognition Workshop and HICSS-36. Kona, Hawaii.

Miller, S.L. (1998). Using ACT-R to model dynamic team behavior. Presented at the Fifth Annual ACT-R Summer School, Carnegie Mellon University, Pittsburgh, PA.

Miller, S.L., Adelman, L., Henderson, E.D., Schoelles, M., and Yeo, C. (2000). *Team decision-making strategies: Implications for designing the interface in complex tasks*. Presented at the Annual Meeting of the International Ergonomics Association and Human Factors and Ergonomics Society, San Diego.

Pozos, R. (2000). *Physiological correlates of Cognitive Tasks*. Presentation at the University of California, San Diego.

Pozos, R. S. (1999, July). *EMG workshop: Correlation of EMG signals with Sequential Finger Movement.* Presented at the International Congress of Biomechanics in Calgary, Canada.

Pozos, R. S. (1999, June). *Finger Force Measurments: A Bioenginneering Approach.* Presented at the University of Minnesota: School of Medicine Invited Speaker Series.

Pozos, R. S. (1999, August). *EMG and Force: Measurements of Fatigue.* Presented at the UCSD Orthopaedic Department.

Schoelles, M. J., & Gray, W. D. (2000). *Argus Prime: Modeling emergent microstrategies in a complex simulated task environment.* Paper presented at the Third International Conference on Cognitive Modeling, Groningen, NL.

Schoelles, M. J., & Gray, W. D. (2000). *Argus Prime: Empirical test of the Argus Prime ACT-R/PM model at the unit task level,* Paper presented at the Seventh ACT-R Workshop. Pittsburgh, PA: Carnegie Mellon University.

Schoelles & Gray (1998, October) Demonstration of Argus Prime. Human Factors and Ergonomics Society 42th Annual Meeting, Chicago.

Stricker, John (August, 2000). Integrating Eye Movements, Visual Memory and Motor Response in a Simple Dynamic Environment. Presented at the 7th annual ACT-R Workshop, Pittsburgh, PA.